

PALASH SETHI

2800 SW Williston Rd, Gainesville, FL 32608
352-709-3301 [◇ palash.sethi@ufl.edu](mailto:palash.sethi@ufl.edu) [◇ palash-sethi.netlify.app](https://palash-sethi.netlify.app)

EDUCATION

- University of Florida** 08/2022 - present
- PhD - Computational Biology
Advisor: Dr Juannan Zhou
Ongoing project: Inferring genotype-phenotype maps using deep learning models
- Birla Institute of Technology and Science-Pilani, Pilani, India** 08/2014 - 06/2019
- M.Sc.(Hons.) Biological Sciences
- B.E.(Hons.) Electronics and Instrumentation Engineering
Thesis: Prediction of CRISPR regulated self-target genes in Salmonella Typhimurium

RESEARCH INTERESTS

- Protein representation and functional optimization
- Functional prediction of proteins from sequential and structural inputs
- Genomic architecture using deep learning

PATENTS

- [1] **Palash S.**, Pranjal B., Anurag S., Shailesh K. CO-GWAS: A co-occurrence graph based GWAS System (Filed, August 2021)¹
- [2] **Palash S.**, Shailesh K. Method and System for learning behaviour of highly complex and non-linear systems - Complexity Based Oracle Sampling (Filed, July 2021)¹
- [3] **Palash S.**, Shailesh K. Non-Linear Constraint Optimization using Neural Networks (Filed, July 2021)¹

WORK EXPERIENCE

- Data Science Team Lead, Hydrocarbons Team, Reliance Jio, India** 02/2021 - 7/2022
Data Scientist, Machine Learning Research Team, Reliance Jio, India 05/2019 - 01/2021
Mentors: Dr. Shailesh Kumar, Dr. Chandra Sarvanan

Co-Occurrence Based Genome Wide Association Study (CO-GWAS) System

- Developed efficient system for genotype to phenotype mapping and identifying CRISPR targets.
- Created custom tensor data-structures to compress (72 TB to 450 GB) and store genome Big Dataset.
- Quantified the relevance of singleton locus to predict phenotype.
- Identified pairwise loci that are together more relevant than singleton locus using a consistency metric formulated by calculating co-occurrence probabilities.
- Identified locus communities by implementing a grow-shrink algorithm on consistency graph. Used concepts from applied graph theory, co-occurrence, community detection and ensemble of models.
- Implemented pipeline for fast and parallelized GWAS across full 3000 rice genomes to identify "Single Nucleotide Variations" achieving 87.3 percent mapping accuracy (running time per exp. : 14 mins.)

Digital Twin for Crude Oil Refineries

- Modeled kinetic equations for refineries using shallow neural networks, providing accurate predictions and faster optimization.
- Formulated non-linear objective functions for constrained optimization of refinery control parameters.

¹drafts and white paper will be shared upon request

- Utilized convex optimization techniques to maximize profits while being constrained by market prices.
- Transformation of POC to a SaaS platform.

Complexity Based Sampling Technique

- Devised an algorithm for efficient sampling by space partitioning over a black-box oracle model to generate training data for ANNs (Decision Trees, Graph-based Algorithms).

RESEARCH EXPERIENCE

Research Affiliate, SciWhy Lab, SCIS, JNU-New Delhi² 04/2021 - 12/2021
Undergraduate Researcher, SciWhy Lab, SCIS, JNU-New Delhi 05/2017 - 07/2017
Advisor: Prof. Shandar Ahmad

Prediction of DNA and RNA binding proteins from sequence

- Developing a motif based Convolutional Neural Network (CNN) using protein PSSM matrix as input.
- Understanding the role of annotations (gene ontology and Pfam) for generating training data.
- Evaluating transfer learning amongst species for DNA/RNA binding protein prediction.

Compression Tool for Gene Sequence

- Implemented burrows-wheeler transform and huffman encoding to compress genome wide sequence data and achieved a compression rate of 0.25X on Mycobacterium Tuberculosis bacteria.
- Analysed spoligotyping in Mycobacterium Tuberculosis bacteria computationally and developed deep learning models for spoligotyping prediction in multiple species.

Research Affiliate, BITS-Pilani, India³ 02/2021 - 12/2021
Undergraduate Researcher, BITS-Pilani, India 01/2019 - 05/2019
Advisor: Dr. Sandhya Marathe

Prediction of CRISPR regulated self-target genes in Salmonella Typhimurium

- Identified self-target genes where translation was regulated by CRISPR1 and CRISPR2 system in Salmonella Typhimurium str 14028s.
- Calculated thermodynamic features - free energy of the mRNA-spacer RNA duplex, and site accessibility energy of the potential mRNA target.
- Analysed biological pathways of the predicted regulated gene set responsible for biofilm production.
- Explored potential Protospacer Adjacent Motif (PAM) sequence from the shortlisted target genes
- Extracted CRISPR spacer targets (plasmids, phages and IMGVR database) and analysed it's conservation across Salmonella serovars of varied host specificity.

Project Assistant, Indian Institute of Science, Bangalore, India 06/2018 - 12/2018
Advisors: Dr. Abhay Sharma, Prof. Raghu Krishnapuram at Robert Bosch Centre for Cyber-Physical Systems @ IISc

Vehicular Traffic Parameters Estimation

- Coded python pipeline to estimate vehicle count and speed using object detection and tracking tools.
- Tracking implemented using Kalman filters resulted in mAP metric of 88 percent.
- Participated in manual data labelling, negotiating with stakeholders for GPU systems, getting cameras fixed on roads, training YOLO models for detection and DEEPSort for tracking vehicles.
- Pipeline adopted by Government of Karnataka at Electronic City, Bangalore for traffic control.

Undergraduate Researcher, BITS-Pilani, India 08/2017 - 12/2018
Advisors: Prof. Shibasish Chowdhury

Prediction and Modeling of miRNA-mRNA interactions

- Parsed miRBase.org using selenium webdriver and BeautifulSoup for mRNA-miRNA interaction data.

²School of Computational and Integrative Sciences (SCIS), Jawaharlal Nehru University (JNU), India

³Birla Institute of Technology and Science-Pilani, India

- Developed a command line tool which generates a quantitative score for mRNA-miRNA binding based on sequential and thermodynamic properties of the complex.
- Trained ML models on the extracted biological features for predicting mRNA-miRNA interactions.

ACADEMIC PROJECTS

Modelling animal invasive species distribution

01/2018 - 05/2018

Advisor: Dr. Sandhya Marathe

- Validated performance of simple models (decision trees, naive bayes), complex models (gradient boosting machines, neural networks) and popular models (Maxent) for modeling invasive species distribution.
- Analyzed the influence of using pseudo-negative samples to assess a model.
- Proposed an alternate approach based on generative adversarial networks to model species distribution using presence only data.

Iterative Vessel Segmentation of Fundus Images

08/2017 - 11/2017

Advisor: Prof. K. K. Gupta

- Parsed DRIVE (Digital Retinal Images for Vessel Extraction) database of retinal images.
- Applied line detection algorithm on multiple scales to separate blood vessel structure from the background in a given fundus image. Otsu's method was applied obtaining a threshold to binarize the multi-scale line response.

ACADEMIC SCHOLARSHIPS/FELLOWSHIPS

- Grinter award for recruiting truly exceptional graduate students at Department of Biology, University of Florida (2022-2025)
- MCN scholarship for excellence in academics, Birla Institute of Technology and Science-Pilani, India (2016-2018).

TECHNICAL EXPERTISE

Computational tools	Python, C++, R, MATLAB
Data Analysis	MaxQuant, Perseus, WEKA
Deep Learning	PyTorch
Biology	PCR, Gene Cloning

REFERENCES

Dr. Juannan Zhou
 Assistant Professor
 Dept. of Biology
 University of Florida
 Email: juannan.zhou@ufl.edu

Prof. Sandhya Marathe
 Assistant Professor
 Dept. of Biological Sciences
 Birla Institute of Technology and Science, Pilani
 Email: sandhya.marathe@pilani.bits-pilani.ac.in

Dr. Shailesh Kumar
 Chief Data Scientist
 Reliance Jio-AI Center of Excellence, India
 Email: Shailesh.Kumar@ril.com

Prof. Shandar Ahmad
 Professor
 Dept. of Computational and Integrative Sciences
 Jawaharlal Nehru University, New Delhi
 Email: shandar@jnu.ac.in